[1] S S V S S R S Sarma Adithe[1]

R K Charpe[2]

Deep Learning-Based Medical Image Segmentation using Dual Decoder Recurrent Residual U-Net (DDR2U-Net) Architecture

Journal of Electrical Systems

Abstract: - Medical image segmentation is a crucial step in the diagnosis and treatment planning of various diseases, particularly in radiology and pathology. Traditional segmentation techniques often struggle with the complex, noisy, and heterogeneous nature of medical images. To address these challenges, we propose a novel deep learning-based architecture, Dual Decoder Recurrent Residual U-Net (DDR2U-Net), designed for accurate and efficient medical image segmentation. The proposed model integrates two key innovations firstly, a dual decoder mechanism, enabling the network to extract both high-level semantic and low-level spatial features simultaneously, enhancing segmentation accuracy and resolution; and second, recurrent residual connections, which improve feature learning by incorporating temporal context and ensuring smoother gradient flow during training. The dual decoder pathways allow for better feature representation by disentangling complex anatomical structures, while the recurrent residual connections help the model retain essential spatial information across multiple layers. Extensive experiments on benchmark medical image datasets, including MRI, CT scans, and histopathological images, demonstrate that the DDR2U-Net outperforms existing state-of-the-art architectures in terms of segmentation accuracy, boundary delineation, and robustness to noise. The proposed model shows promise for applications in automatic organ segmentation, tumor detection, and other critical medical imaging tasks. The model was experimentally verified using the publicly available Kvasir-SEG dataset, which gives a better global accuracy, Recall, Precision Score, IoU and Dice compared to the prior works ColonSegNet, UPolySeg and R2UPolyseg. These results show an improvement in accuracy obtained by DDR2UPolySeg

Keywords: Medical Imaging, Segmentation, Convolutional Neural Networks, U-Net, Residual U-Net, Recurrent U-Net, R2U-Net and DDR2U-net.

I. INTRODUCTION

Image processing is frequently discussed in relation to computer vision. The deep learning technique is one that is frequently applied in image processing [1] in the process. The Deep Learning technique has demonstrated its effectiveness in numerous sectors of application [2]. Convolutional neural networks (CNNs) are one of the most often used deep neural network techniques these days [3]. The terms "deep learning" and "deep neural network" apply to multi-layer ANNs and CNNs. CNN has advanced significantly, particularly in tasks pertaining to image processing and vision [4]. Particularly when talking about image processing, this will be extremely helpful for technical advancements in the field of artificial intelligence, which have up till now been a trend. The benefits of image processing itself are numerous and can be used to meet needs, like grouping, prediction, classification, and so forth.

Pixel-by-pixel categorical information can be obtained by semantic segmentation. This job has several real-world applications, including computer-assisted diagnosis, therapeutic planning, pedestrian detection, self-driving cars, and handicap identification. Every image pixel is given a category label by semantic segmentation. Intelligent systems can comprehend spatial locations or make critical decisions with the use of pixel-level semantic information [5]. Deep CNN is frequently utilized in the medical domain to solve biomedical segmentation issues [6]. The segmentation model is one of the recent improvements that CNN has made [7]. The U-Net Convolutional Network is one of the CNN designs used for image segmentation [8]. Since its introduction in 2015, this architecture has changed quickly. For semantic segmentation applications, U-Net is frequently utilized in the medical industry [9]. The encoder-decoder architecture, on which U-Net is built, allows for the very effective use of an encoder-decoder-based deep learning technique to solve a variety of AI application challenges [10]. Using a skip link, U-Net integrates a high-level semantic feature map from a low-level feature map decoder on the encoder [7].

Biomedical signal acquisition, image synthesis from signals, image processing, and picture display for medical diagnosis are all areas of rapidly developing research in the subject of biomedical image processing. Devices for clinical imaging integrate hardware and software. Researchers in the field of biomedical image processing are

[1] [1]Research Scholar, Kalinga University, sarmaadithe99@gmail.com

[2]Professor, Kalinga University

interested in the quantity of medical imaging sensors. Researchers in the field of biomedical image processing are interested in the quantity of medical imaging sensors. A subfield of image classification is medical image classification. Lung cancer is the most deadly type of cancer of all the others. This is due to the fact that non-small cell lung cancer (NSCLC) patients typically receive an advanced diagnosis. The death rate attributable to lung cancer even reached 18.4% of the total fatalities caused by cancer worldwide, and this was determined in 2019 [11]. As a result, lung segmentation is crucial because it effectively determines the lung region on CT scan pictures, which is helpful for treating lung-related illnesses.

Colorectal cancer (CRC) is one of the type of cancer that occurs in the colon or rectum, which are parts of the digestive system. This is the one of the most common cancers worldwide and a significant cause of cancer-related deaths, especially in developed countries. Understanding colorectal cancer, its risk factors, symptoms, and prevention strategies is crucial for early detection and treatment. Colorectal cancer occurs when abnormal cells in lining of the colon or rectum grow uncontrollably, forming a tumor. As time goes on, these cancer cells maybe invade nearby tissues and organs and spread to other parts of the body via a blood system or lymph system (methaaster).

Notably, colorectal cancer (CRC) is the second most deadly kind of cancer in young women and the deadliest in young men (ages 20 to 49). About 10% of new occurrences of colorectal cancer (CRC) occur in people under 50, reflecting the disease's steady rise in frequency among younger persons. The overall decreasing incidence rates since the mid-1980s, which have been mostly attributed to better screening and lifestyle modifications, are in contrast to this increase, which is occurring at a pace of 1% to 2% year.

The risk of colorectal cancer can be considerably reduced by finding polyps early on. Segmenting colonoscopy images can expedite the diagnosis of polyps, however, due to differences in polyp size, shape, and location, precise segmentation is difficult to achieve. Furthermore, the proficiency of the person executing the task has a direct bearing on the accuracy of polyp image segmentation. Fig. 1.4 shows the large intestine, the lower part of the large intestine is the rectum and the upper part is the colon.

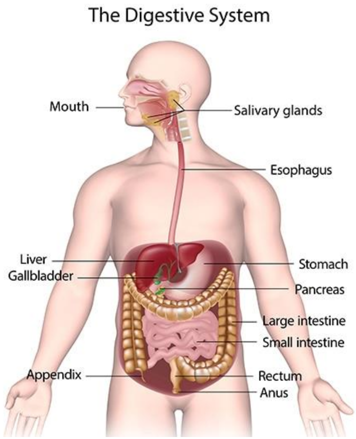

Fig 1: Anatomy of the Gastrointestinal System

Based on their shape, polyps are classified into two types: (1) sessile (flat) and (2) pedunculated (stalk-bearing). Sessile polyps lie flat on the surface of the inner wall of the colon or rectum and are therefore difficult to detect. Pendulum polyps resemble mushrooms and are attached to the lining of the colon by a long, thin stalk. Based on pathological characteristics, polyps are classified into five categories: adenomatous, hyperplastic, serrated, inflammatory, and villous. Adenomatous polyps have a tubular structure and are the most common form of polyp. Over time, benign adenomatous polyps can turn into malignant polyps. Therefore, identifying and removing benign adenoma tumors can halt the development of colorectal cancer.

Hyperplastic polyps are small in size and have a low risk of developing into cancer. Serrated polyps can develop into cancer depending on their size and location. Small, jagged polyps are called hyperplasia. Larger, serrated polyps in the upper part of the colon are usually flat (sessile), harder to detect, and precancerous. Inflammatory polyps are also called pseudopolyps because they are not true polyps but are formed due to chronic inflammation of the colon or rectum. Inflammatory polyps are often benign and do not pose a risk of developing colorectal cancer.

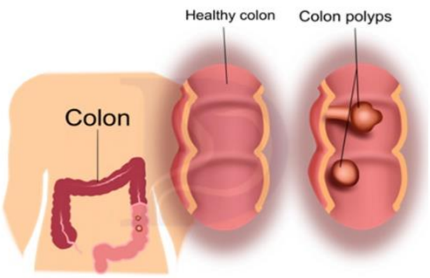

Figure 2: Visualization of the colon and polyps

Inflammatory polyps are also called pseudopolyps because they are not true polyps but are formed due to chronic inflammation of the colon or rectum. Inflammatory polyps are often benign and do not pose a risk of developing colorectal cancer. The chorus glioma, also known as urinary tube glioma, has a serious risk of cancer. They usually stick, making it more difficult to detect and delete. Doctors offer regular screening to detect and eliminate polyps as soon as possible. Discan and mild polyps are probably removed by surgery for new forecasts. There are three types of exclusion selection. The most common type is withdrawal using pliers or wire cycle. This process is called polypectomy. If a polyp is too large to eliminate with this method, a liquid can be injected under it to lift and isolate the polyp of the surrounding tissues so that it can be removed. The second option is minimally invasive surgery. Polyps that are too large or cannot be safely removed during screening are usually removed surgically, which is often done by inserting an instrument called a laparoscope into the abdomen to remove the diseased part of the intestine. The last option is a total proctocolectomy, in which the colon and rectum are permanently removed to prevent the development of potentially fatal colorectal cancer.

Colonography:

Colonoscopy is a medical procedure used to examine the inside of the large intestine (colon) and rectum, which are important parts of the digestive system. The procedure is widely considered one of the most effective tools for detecting and preventing colorectal cancer, as well as diagnosing other digestive diseases. Coloscopy includes the use of a long flexible tube called Coloscope, in which there is a small video camera attached to its tip. The coloscope is inserted through the rectum and carefully goes all the length of the colon. The camera transmits real-time images to a monitor, allowing your doctor to carefully examine the lining of the colon and rectum for abnormalities such as polyps, tumors or areas of inflammation.

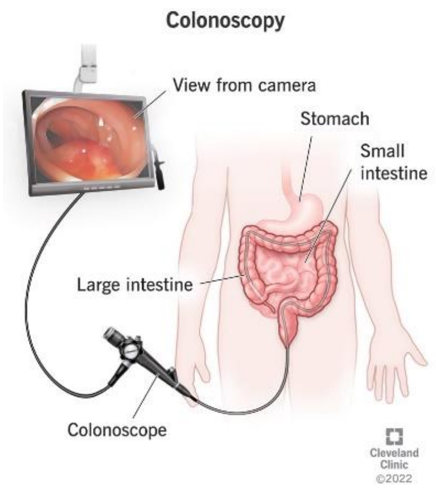

Fig 3: Colonoscopy Screening Test

Fig 4: Various Polyp images captured using Colonoscopy

2. LITERATURE REVIEW

Polyps are an essential sign of early colon cancer, so the main purpose of the examination is to detect them as early as possible to improve patient survival rates. Automatic detection and localization of polyps in video frames of gastrointestinal endoscopy can help reduce missed and false detections in manual manipulation, improve detection quality and efficiency, and have positive implications for early detection of pre-cancerous lesions. Within the realm of medical image analysis, directional radiotherapy, image-guided interventional diagnosis and treatment, and other procedures can all benefit from the application of medical image segmentation. Presently, the most effective means of polyp detection and colorectal cancer prevention is optical colonoscopy. However, colon screening demands considerable time and relies heavily on the operator's skill. Hence, there's a need to create a computer-aided diagnosis (CAD) technique to autonomously segment polyps in colonoscopy images. Within image analysis, the extensive adoption of deep learning techniques has outperformed traditional methods significantly. Extracting polyp features from training datasets involves fitting a non-convex network with non-linear neuron objectives. The deep architecture inherent in neural networks can streamline these chosen features for future classification. CNNs, owing to their impressive achievements, have garnered substantial popularity and widespread utilization in medical image analysis.

Q. Nguyen and S.-W. Lee [2018] has proposed Colorectal Segmentation Using Multiple Encoder-Decoder Network in Colonoscopy Images. The authors tried to improve the segmentation accuracy of colorectal structures in colonoscopy images, which is critical for the detection and diagnosis of colorectal diseases. For this the authors proposed a multiple encoder-decoder network that leverages the strengths of deep learning architectures to effectively segment complex anatomical structures. This architecture enables the model to capture both local and global features, enhancing segmentation performance. The model is trained and evaluated using a dataset of colonoscopy images, which includes various types of colorectal structures. The authors emphasize the importance of high-quality annotated data for training deep learning models effectively.

Z. Zhou, et. al., [2018] proposed a A Nested U-Net Architecture for Medical Image Segmentation. The authors addresses the limitations of existing U-Net architectures in medical image segmentation by proposing UNet++, which incorporates a nested structure to enhance feature extraction and improve segmentation accuracy. UNet++ features a series of nested skip pathways that connect the encoder and decoder sub-networks. This design reduces the semantic gap between the feature maps generated by the encoder and decoder, facilitating better learning and feature representation.

D. Jha et al., [2021] proposed NanoNet, a novel deep learning architecture for real-time segmentation of polyps in video capsule endoscopy and colonoscopy images. NanoNet follows an encoder-decoder approach, using a pre-trained MobileNetV2 model as the encoder to capture contextual information from the input images. The decoder consists of modified residual blocks to generate the final segmentation output using the features extracted by the encoder. The authors used the KvasirCapsule-SEG dataset for training and evaluating NanoNet. The dataset contains images from both video capsule endoscopy and colonoscopy. This NanoNet represents a significant advancement in real-time polyp segmentation for endoscopic imaging, combining an efficient encoder-decoder architecture with pre-trained models to achieve better accuracy and speed.

An architecture designed to enhance the segmentation of polyps in colonoscopy images has presented by N.K. Tomar, D. Jha, S. Ali, et al., [2020]. This work has developed an automated method for accurately segmenting colorectal polyps, which is crucial for improving diagnosis and treatment planning. The DDANet is a fully convolutional network that features a single encoder and dual decoders. The encoder processes the input RGB image, while the dual decoders generate both a segmentation mask and a grayscale version of the input image.

T. Mahmud et al., [2021] has proposed a PolypSegNet which is Modified Encoder-Decoder Architecture for Automated Polyp Segmentation from Colonoscopy Images. The authors presented automated system for accurately segmenting polyp regions in colonoscopy images to facilitate faster and more reliable diagnosis of colorectal cancer. PolypSegNet is a modified encoder-decoder architecture that enhances traditional segmentation methods.

S. Y. Park et al., [2012] has presented a Colon Video Analysis Framework for Polyp Detection. In this study the authors developed an automated video analysis framework that improves the detection of colonic polyps, which is essential for early diagnosis and treatment of colorectal cancer. This framework integrates both spatial frame-based analysis and temporal video analysis. This dual approach allows the system to analyze time-course image sequences effectively, enhancing the detection capability by considering both individual frames and their temporal relationships.

M. Misawa et al., [2021] has proposed Development of a Computer-Aided Detection System for Colonoscopy and a Publicly Accessible Large Colonoscopy Video Database. In this study the authors developed a computer-aided detection (CAD) system to assist endoscopists in identifying polyps during colonoscopy, thereby reducing the likelihood of missed adenomas and improving diagnostic accuracy. The authors constructed a deep learning-based AI system using a dataset of 56,668 independent colonoscopy images collected from five medical centers. This dataset served as the foundation for training the AI model.

Z. Guo et al., [2021] has presented Polyp Detection Algorithm Can Detect Small Polyps: Ex Vivo Reading Test Compared with Endoscopists. In their study, the authors aimed to assess the effectiveness of a computer-aided detection (CAD) algorithm in detecting small polyps, which are often missed by endoscopists during routine examinations. The authors conducted an ex vivo reading test using colonoscopy videos that included various polyp sizes. The performance of the CAD algorithm was compared against the detection capabilities of experienced endoscopists. The dataset used for testing included a range of polyp sizes, with a particular focus on smaller polyps, which are typically more challenging to detect.

Mohammed Ahmed et al., [2020] has proposed PS-DeVCEM: Pathology-sensitive deep learning model for video capsule endoscopy based on weakly labelled data. The authors introduced PS-DeVCEM, a pathology-sensitive deep learning model designed for frame-level anomaly detection and multi-label classification of various colon conditions in VCE. The model is trained using weakly labeled data, where only video-level labels are provided instead of precise frame-level annotations. This reduces the burden of manual labeling. PS-DeVCEM employs a multiple instance learning (MIL) approach to handle the weakly labeled data. It learns to predict video-level labels by aggregating frame-level predictions. The model is evaluated on a dataset of 300 VCE videos, demonstrating its ability to detect and classify abnormalities such as polyps, ulcerative colitis, and Crohn's disease.

J. G.-B. Puyal et al., [2022] has proposed Polyp Detection on Video Colonoscopy Using a Hybrid 2D/3D CNN. In this work the authors aimed to enhance the accuracy of polyp detection in colonoscopy videos by developing a hybrid convolutional neural network (CNN) that integrates both 2D and 3D convolutional layers. This hybrid model leverages 2D CNNs to extract spatial features from individual frames and 3D CNNs to capture temporal information across multiple frames. This combination allows the model to effectively analyze the dynamic nature of colonoscopy videos. The model is trained and evaluated on a comprehensive dataset of colonoscopy videos, which includes a variety of polyp types and sizes.

M. Liu, J. Jiang, and Z. Wang [2019] has proposed Colonic Polyp Detection in Endoscopic Videos with Single Shot Detection Based Deep Convolutional Neural Network. In this work, the authors developed a system for real-time detection of colonic polyps during endoscopic procedures, addressing the challenge of missed polyps by endoscopists. In this work, the authors proposed a Single Shot Detection (SSD) framework, which integrates a deep convolutional neural network (CNN) for detecting polyps in video frames. This approach allows for simultaneous object detection and classification, enabling real-time analysis. The model is trained on a dataset of endoscopic images that includes various types of polyps, ensuring that the system can generalize well to different polyp appearances.

3. DATASETS

Knowing the importance of supervised learning techniques for automated disease detection, different group of researchers and organizations have developed various datasets by collecting medical images from radiology centers, hospitals, and cancer research institutes. Many datasets in the reviewed papers have been kept private due to privacy concerns, and reproducing their results is hence not possible. and their descriptions, including the main reference, imaging technique, number of images, and the URL. It can be seen that there are no large scale datasets available to date in the medical imaging field like ImageNet, which often poses challenges in designing effective CNN models.

In addition to the publicly available dataset, we colleborated with Aichi Medical University, Nagakute, Japan and generated a database under the supervision of Kunio Kasugai at the Department of Gastroenterology. The dataset consists of colonoscopy videos recorded with Narrowband Imaging (NBI) and White Light (WL). The expert has selected the frames which contain visible polyps. For stage 1 classification, 900 WLimagesareavailablewhere400images contain polyps, and the rest of the images does not contain any polyps. For stage 2 classification, 400 NBIimageswith275non-neoplastic images and 190 WL images, where 125 are neoplastic, and 107 are non-neoplastic, are available

Furthermore, keep the concern of reproducibility and verification, we use public dataset as well in our work. The details of each of the public dataset used in this work are,

Kvasir-SEG: The Kvasir-SEG dataset is a valuable resource in the field of medical imaging, specifically designed for the detection and segmentation of gastrointestinal polyps in colonoscopy images. The Kvasir-SEG dataset consists of high-quality images from GI endoscopy procedures, specifically annotated for the task of segmentation. Each image is paired with ground truth segmentation masks that delineate the regions of interest, such as polyps and other GI lesions. The Kvasir-SEG dataset contains 1,000 polyp images along with their corresponding segmentation masks. These images were sourced from the larger Kvasir dataset, specifically tailored for segmentation tasks. The images in the Kvasir-SEG dataset vary in resolution, ranging from 332 x 487 to 1920 x 1072 pixels. This diversity in resolution allows for the evaluation of algorithms across different image qualities and sizes. Figure 4 shows the sample polyp images and corresponding ground truth images of Kvasir-SEG dataset.

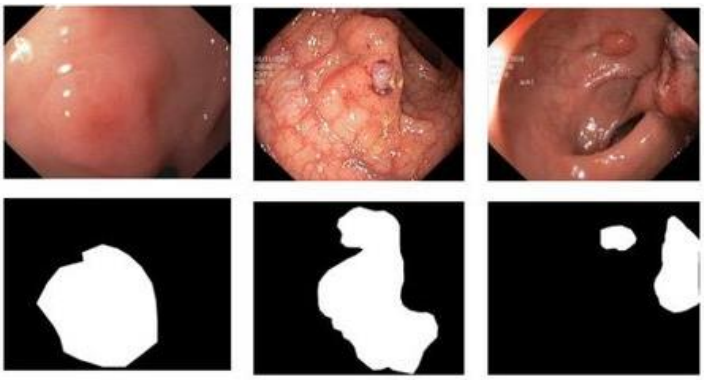

Fig 5. Examples of content of Kvasir-SEG dataset. First row shows original images whereas second row shows corresponding ground truth

4. DUAL-DECODER-R2U-NET ARCHITECTURES

Dual-Decoder-R2U-Net is an advanced variation of the U-Net architecture that combines the Dual Decoder and R2U-Net features.

- **Dual Decoder**: It includes two independent decoder paths for handling multiple outputs.
- **R2U-Net**: It integrates the Recurrent Residual Convolutional Neural Network (RRCNN) concept, which combines residual connections and recurrent convolutional layers, improving feature representation.

Components of Dual-Decoder-R2U-Net:

i. **R2U-Net Structure:**
   - **Residual Connections**: Borrowed from ResNet, residual connections help combat the vanishing gradient problem by allowing the network to learn identity mappings more easily, which improves the training of deeper networks.
   - **Recurrent Convolutional Layers (RCLs)**: These layers apply the same convolutional operation multiple times, allowing the network to capture temporal or contextual dependencies in spatial features. This improves the ability to extract complex patterns from input images.

ii. **Dual Decoder:**
   - After the shared encoder and R2U-Net bottleneck, the network has two separate decoder branches, each tasked with generating a distinct output from the same input. These outputs can be different segmentation maps or other forms of image understanding.
   - Each decoder consists of upsampling operations (typically transposed convolutions) and is complemented by skip connections from corresponding encoder layers. The skip connections allow the model to use both high-level semantic information and fine-grained details to generate precise segmentations.

iii. **Encoder (Downsampling Path):**
   - The encoder consists of convolutional layers combined with max-pooling operations, progressively reducing the spatial dimensions of the input while capturing important features.
   - Each encoder block includes Recurrent Convolutional Layers and Residual Blocks to enhance feature extraction.

iv. **Bottleneck Layer:**
   - The bottleneck is the lowest layer of the network, which compresses the learned features before they are passed to the decoders.

v. **Dual Decoders (Upsampling Path)**:
   - Each decoder path reconstructs its output by progressively upsampling the bottleneck features and concatenating them with the corresponding encoder features via skip connections. This ensures that both high-level and detailed features are retained.

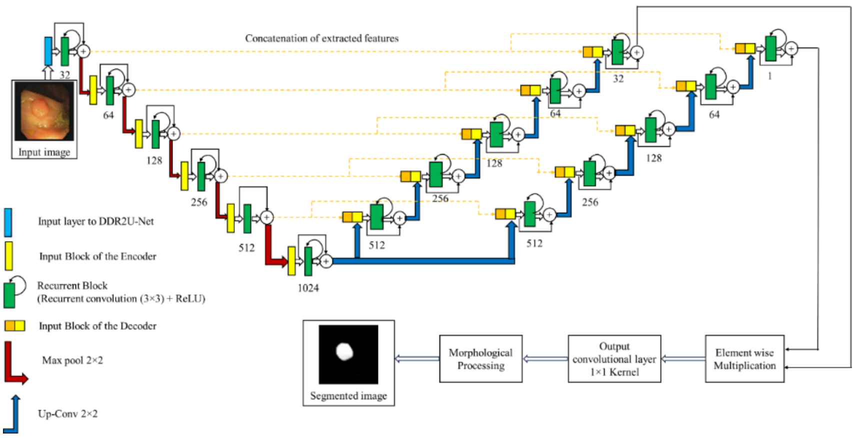

Fig 6. Network architecture of the Proposed DDR2U-Net model

5. SIMULATION RESULTS

To demonstrate the performance of the Dual decoder based R2U-Net model, we have tested on different medical images from different datasets such as Kvasir-SEG. These particularly focused on the segmentation of polyps in colonoscopy images.

The network was evaluated using various parameters, such as global accuracy (GA), dice coefficient (DC), intersection over union (IoU), recall (R), and precision (P). The DDR2UPolySeg model was trained using different hyper parameters. Here, the global accuracy represents the proportion of correct predictions. The global accuracy is calculated using Equation (1). The intersection over union, also known as the Jaccard index, shows the proportion of overlap between the predicted value and the ground truth mask (represented in Equation (2)). The dice coefficient is quite similar to the IoU, but it double counts the intersection, as shown in Equation (3). Precision signifies the purity of a positive detection compared with the ground truth, whereas recall signifies the completeness of a positive detection compared with ground truth. Precision and recall can be evaluated using Equation (4) and Equation (5), respectively. Each of the parameters was evaluated by taking into account the true-positive (T_p), true-negative (T_n), false-positive (F_p), and false-negative (F_n) rates.

$$GA = \frac{T_p+T_n}{T_p+T_n+F_p+F_n} \quad (1)$$

$$IoU = \frac{T_p}{T_p+F_p+F_n} \quad (2)$$

$$DC = \frac{2.T_p}{2.T_p+F_p+F_n} \quad (3)$$

$$P = \frac{T_p}{T_p+F_p} \quad (4)$$

$$R = \frac{T_p}{T_p+F_n} \quad (5)$$

The proposed DDR2UPolySeg model was trained on a system with Intel Core 2.60 GHz i7 CPU running Windows 10 with 16 GB RAM, NVIDIA GeForce GTX 1650 GPU. All of the experiments were performed using Jupyter Notebook.

Figure 6 illustrates an overlay of the final segmented image along with the ground truth image for Kvasir-SEG dataset. Table 1 presents the calculated values of the evaluation parameters for DDR2UPolySeg compared with UPolySeg. It is observed from the evaluation matrix (Table 5.3) that the DDR2UPolySeg model performed better than ColonSegNet and UPolySeg for Kvasir-SEG dataset.

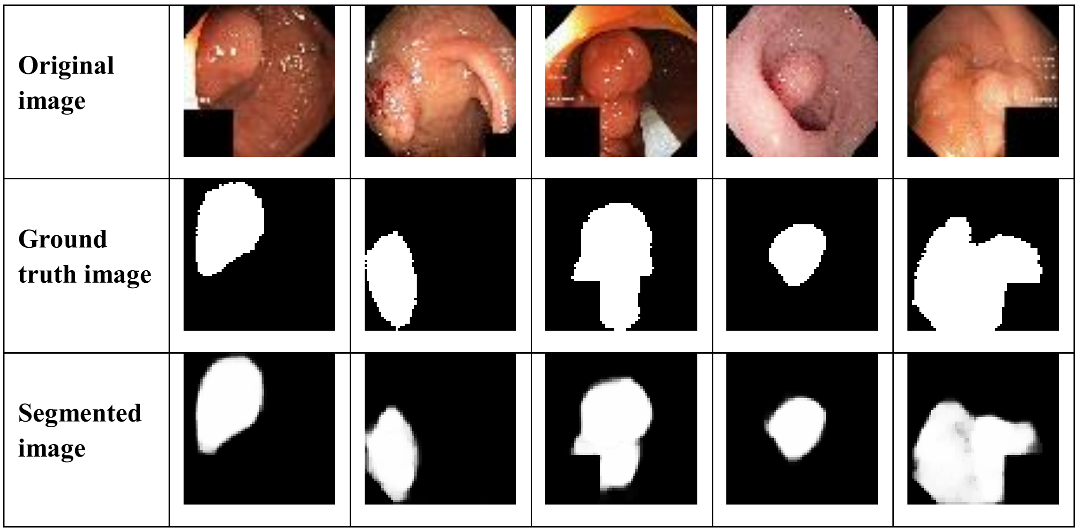

Figure 7. The final segmented image along with the ground truth image for CVC-Colon DB dataset. (a) Original image, (b) ground truth image, (c) segmented image for CVC-Colon DB dataset

Table 1. Performance evaluation of DDR2UPolySeg with ColonSegNet and UPolySeg for CVC- Kvasir-SEG dataset

Model	**Accuracy**	**Recall**	**Precision Score**	**IoU**	**Dice**
ColonSegNet	0.8862	0.7952	0.8145	0.8223	0.8164
UPolySeg	0.9128	0.8654	0.8659	0.8514	0.8723
DDR2UPolySeg	0.9379	0.8941	0.8734	0.8553	0.8918

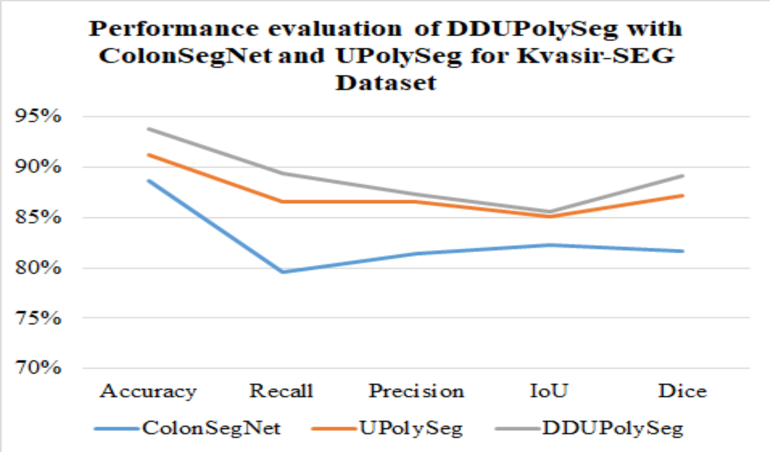

Fig 8. Performance evaluation of DDR2UPolySeg with ColonSegNet and UPolySeg for Kvasir-SEG dataset

6. CONCLUSION

In this paper, we introduced the Dual Decoder Recurrent Residual U-Net (DDR2U-Net), a novel deep learning architecture designed to address the challenges inherent in medical image segmentation. By incorporating a dual decoder mechanism and recurrent residual connections, DDR2U-Net is capable of capturing both high-level semantic features and low-level spatial details, while effectively modeling long-range dependencies within medical images. These innovations significantly improve segmentation accuracy, particularly for complex anatomical structures and boundary delineation. Our extensive evaluations on multiple medical image datasets, demonstrate that DDR2U-Net outperforms several state-of-the-art models in terms of segmentation accuracy, robustness, and efficiency. The dual decoder structure enables the network to better retain fine-grained details, while the recurrent residual connections enhance the model's ability to learn contextual information across deeper layers. These features make DDR2U-Net particularly effective for tasks such as tumor detection, organ segmentation, and the delineation of irregular structures in noisy medical images. The results highlight DDR2U-Net's potential as a valuable tool for clinical applications, offering improvements in both performance and reliability over existing automated segmentation techniques. By reducing reliance on manual annotation and improving segmentation precision, the proposed model could significantly aid clinicians in diagnosis, treatment planning, and follow-up assessments

References:

[1] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 7068349, 1-14. doi:10.1155/2018/7068349.

[2] Soydaner, D. (2020). A Comparison of Optimization Algorithms for Deep Learning. International Journal of Pattern Recognition and Artificial Intelligence, 34(13), 1–26. doi:10.1142/S0218001420520138.

[3] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2018). Understanding of a convolutional neural network. Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017, 1–6. doi:10.1109/ICEngTechnol.2017.8308186.

[4] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8), 5455–5516. doi:10.1007/s10462-020-09825-6.

[5] Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. Neurocomputing, 406, 302– 321. doi:10.1016/j.neucom.2019.11.118.

[6] Salehi, S.S.M., Erdogmus, D., Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. Machine Learning in Medical Imaging. MLMI 2017, Lecture Notes in Computer Science, 10541, Springer, Cham, Switzerland. doi:10.1007/978-3-319-67389-9_44.

[7] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., & Wu, J. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp40776.2020.9053405.

[8] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015, Lecture Notes in Computer Science, 9351. Springer, Cham, Switzerland. doi:10.1007/978-3-319-24574-4_28.

[9] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., & Maier-Hein, K. H. (2019). Abstract: nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. Bildverarbeitung Für Die Medizin 2019, 22–22, Springer Vieweg, Wiesbaden, Germany. doi:10.1007/978-3-658-25326-4_7.

[10] Yalçın, S., & Vural, H. (2022). Brain stroke classification and segmentation using encoder-decoder based deep convolutional neural networks. Computers in Biology and Medicine, 105941, 149. doi:10.1016/j.compbiomed.2022.105941.

[11] Nair, R. R., & Singh, T. (2019). Multi-sensor medical image fusion using pyramid-based DWT: A multi-resolution approach. IET Image Processing, 13(9), 1447–1459. doi:10.1049/iet-ipr.2018.6556.

[12] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[13] S. K. Ghosh and A. Ghosh, “A novel intuitionistic fuzzy soft set based colonogram enhancement for polyps localization,” International Journal of Imaging Systems and Technology, vol. 31, no. 3, pp. 1486–1502, 2021.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[15] D. Jha et al., “Kvasir-SEG: A segmented polyp dataset,” in International Conference on Multimedia Modeling, Springer, 2020, pp. 451–462

[16] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, “Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” Computerized Medical Imaging and Graphics, vol. 43, pp. 99– 111, 2015

[17] H. Borgli et al., “Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy,” Scientific Data, vol. 7, no. 1, pp. 1–14, 2020

[18] S. A. Hicks et al., “The endotect 2020 challenge: Evaluation and comparison of classi- fication, segmentation and inference time for endoscopy,” in International Conference on Pattern Recognition, Springer, 2021, pp. 263–274

[19] X. Yang, L. Yu, S. Li, et al., “Towards automated semantic segmentation in prenatal volumetric ultrasound,” IEEE Transactions on Medical Imaging, vol. 38, no. 1, pp. 180– 193, 2018

[20] J. Chmelik, R. Jakubicek, P. Walek, et al., “Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3d ct data,” Medical Image Analysis, vol. 49, pp. 76–88, 2018

[21] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." Advances in Neural Information Processing Systems. 2017.

[22] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." arXiv preprint arXiv:1511.00561(2015).

[23] W. H. Organization et al., “Global health estimates 2015: Deaths by cause, age, sex, by country and by region, 2000–2015,” Geneva: WHO, p. 2016, 2016.

[24] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, et al., “Cancer statistics, 2021,” Ca Cancer J Clin, vol. 71, no. 1, pp. 7–33, 2021.

[25] T. Shen and Y. Wang, “Medical image segmentation based on improved watershed algo- rithm,” in 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2018, pp. 1695–1698

[26] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431- 3440).

[27] Wang, Naiyan, et al. "Transferring rich feature hierarchies for robust visual tracking." arXiv preprint arXiv: 1501.04587 (2015).

[28] Mao, Junhua, et al. "Deep captioning with multimodal recurrent neural networks (m-rnn)." arXiv preprint arXiv: 1412.6632 (2014)

[29] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv: 1409.1556 (2014).

[30] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[31] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[32] Huang, Gao, et al. "Densely connected convolutional networks." arXiv preprint arXiv:1608.06993 (2016).

[33] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, “The ever-increasing importance of cancer as a leading cause of premature death worldwide,” Cancer, vol. 127, no. 16, pp. 3029–3030, 2021.

[34] H. Sung, J. Ferlay, R. L. Siegel, et al., “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: a cancer journal for clinicians, vol. 71, no. 3, pp. 209–249, 2021.

[35] Y. Jiang et al., “Global pattern and trends of colorectal cancer survival: A systematic review of population-based registration data,” Cancer Biology & Medicine, vol. 19, no. 2, p. 175, 2022.

[36] M. Gao et al., “Comprehensive analyses of correlation and survival reveal informative lncrna prognostic signatures in colon cancer,” World Journal of Surgical Oncology, vol. 19, no. 1, pp. 1–15, 2021.

[37] B. Levin et al., “Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the american cancer society, the us multi-society task force on colorectal cancer, and the american college of radiology,” Gastroenterology, vol. 134, no. 5, pp. 1570–1595, 2008.